Understanding the behavior of foodborne pathogens in the food chain: new information for risk assessment analysis

UNIVERSITÀ DEGLI STUDI DI TORINO

Understanding the behavior of foodborne pathogens in the food chain: new information for risk assessment analysis

Kalliopi Rantsiou[1]*, Marios Mataragas[2], Lene Jespersen[3], Luca Cocolin[1]

[1]DIVAPRA, Faculty of Agriculture, University of Turin, Italy

[2]Laboratory of Food Quality Control and Hygiene, Department of Food Science and Technology, Agricultural University of Athens, Greece

[3]Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Denmark

*Address for correspondence: Via Leonardo da Vinci 44, 10095 Grugliasco, Torino, Italy. Tel. 0039/011/670-8553, Fax 0039/011/670-8549, email: kalliopi.rantsiou@unito.it

Abstract

In recent years and with the significant advancements in instrumentation for molecular biology methods, the focus of food microbiologists, dealing with pathogenic microorganisms in foods, is shifting. Scientists specifically aim at elucidating the effect that the food composition, as well as the commonly employed preservation/storage techniques throughout the food chain, have on the virulence of pathogens. Quantitative PCR and microarrays are, nowadays, powerful tools used for such determinations. The application of these approaches for the determination of the gene expression *in situ*, is a new field of research for food microbiologists and provides new information regarding virulence potential of foodborne pathogens.

Introduction

Despite significant knowledge acquisition regarding food safety, control of foodborne pathogens throughout the food chain remains a challenge for food producers, authorities and consumers. World-wide, various important efforts have been undertaken in order to reduce the incidence of diseases related to the consumption of food, however, so far limited success has been obtained (Skovgaard, 2007). Food production is a dynamic sector, where the consumer's needs are of primary importance. For this reason, the scenario in which food safety has to be guaranteed is often subjected to changes, which affect also the fitness and the behavior of the pathogenic microorganisms. A relevant example is the case of *Listeria monocytogenes*, a psycrotrophic bacterium that emerged as a foodborne

pathogen after establishment of refrigeration throughout the food chain, in order to prolong the shelf-life of foodstuffs maintaining the aspect of freshness requested by the consumers.

Authorities have addressed the issue of pathogenic microorganisms in foods by imposing legislations that either consider limits for their presence or do not tolerate the presence of even one pathogenic cell in the foodstuff analyzed. Thereby it is evident how the food safety aspect is nowadays addressed by the means of numbers. Official analysis is done by traditional microbiological testing, approaches able to enumerate the microorganisms in food, even if it has been repeatedly demonstrated that methods that rely on cultivation of the cells, often fail detection, especially if the pathogen is present in a stressed or injured state. Moreover, the time needed to retrieve the results on the presence/absence of a foodborne pathogen is not appropriate to the time constraints of the modern food industry. Also, the recent risk analysis approaches introduced in the food sector, suffer from the limitation of only considering numbers of viable microorganisms. Based on these considerations there is the need for alternative ways to ensure consumer's health protection.

Nowadays, the field of biotechnology, with the recent advancements in the nucleic acid analyses, is offering a number of choices that can be used. Considering food safety and foodborne pathogens, one aspect that can be explored is the behavior of the microorganisms in the food matrices. This approach takes into consideration the expression of specific traits, namely virulence and stress responses, *in situ*. From studies conducted so far, it has been shown that within species of pathogenic bacteria, strain

heterogeneity in virulence potential exists. Furthermore, environmental conditions that pathogenic microorganisms encounter in foods influence their stress response capabilities, enhance survival and possibly overall virulence potential. The application of *omics* disciplines such as genomics, transcriptomics, proteomics and metabolomics offer significant potential for advancements by improving the understanding of the virulence determinants of pathogenic bacteria (Yoshida et al. 2001). The outcome of studies targeting gene expression to clarify foodborne pathogen behavior will be useful for risk assessment in order to improve food safety (Figure 1) (Brul 2007).

Methodology

Technological advancements in molecular biology methods have over the last 15 years shifted the interest of microbiologists from the study of a single gene and its products, to more global approaches that produce significant amount of biological data in a single experiment. These technological advancements have resulted in a wealth of publically available *genomic data* through full genome sequencing projects for different microorganisms of interest to food microbiologists, mainly foodborne pathogens and technologically important species. Genomic data give only indications of the *potential* of a given microorganism in terms of metabolic activities, survival in different conditions, virulence, stress response, which however may never be expressed. For this reason, scientists nowadays are focusing not only on the generation of new genomic data but also on their exploitation for the understanding of the

true capabilities (for example metabolic activities or virulence expression) of microorganisms in different environmental conditions through the application of *transcriptomics*.

Transcriptomics enable the analysis of the RNA transcripts produced by the genotype at a given time and provide a link between the genome, the proteome and the cellular phenotype. Through this approach, a better understanding of the molecular basis of virulence could be gained and a further insight into the complex expression events involved could be achieved. Technologies that are used in transcriptomics are the following: i) *Microarrays* in order to evaluate the gene expression events providing information on the differentially expressed genes (global) and ii) *Reverse Transcriptional Quantitative Polymerase Chain Reaction (RT-qPCR)* in order to quantify (and confirm) the differential expression of most important genes (Figure 2).

Both approaches can be used in order to determine the amount of cDNA, deriving from a specific mRNA molecule and therefore can give information regarding gene expression. In food microbiology, this approach can be employed to study expression of genes that are involved in the survival, stress response and virulence of pathogenic microorganisms and how it is influenced by environmental conditions in the food chain, the food matrix in which microorganisms are found or by changes in common production or preservation techniques.

Analysis of transcriptional data can be divided into two stages (Livak and Schmittgen 2001; Pfaffl 2001; Quackenbush 2002; Causton, Quackenbush & Brazma, 2003; Wilson, Tsykin, Wilkinson & Abbott, 2006): i)

Transformation of the raw data into a gene expression matrix (e.g. data normalization to account for non-biological variability between samples) and ii) Analysis of the gene expression matrix (e.g. Analysis of Variance – ANOVA, clustering, principal component analysis, multidimensional scaling and classification methods for class prediction). During data analysis no means or median values should be used but all replicate measurements should be included in the matrix as separate columns in order to take into consideration information about the variance. The application of these analytical techniques produces multivariate information and leads to an enormous amount of data raising questions such as 'how someone could get relevant information out of the measured data?', 'how someone could represent and display this information?' and 'how someone could get such information into data?' (Forina, Lanteri & Casolino, 2004).

Quantitative PCR, first described in 1992 (Higuchi, Dollinger, Walsh & Griffith, 1992; Higuchi, Fockler, Dollinger & Watson, 1993) is considered as the next generation of PCR techniques, proposed in 1986 by Mullis, Faloona, Scharf, Saiki, Horn and Erlich (1986), the method that revolutionized research at the molecular level for different scientific fields, including food microbiology. While PCR can be used in food microbiology to give a yes or no answer, qPCR allows the monitoring, in real time, of the synthesis of an amplicon and therefore can be used to quantify the amount of a target DNA molecule present in the initial amplification mix. This method can be used in food microbiology, to indirectly determine, through construction of appropriate calibration curves, the concentration (in terms of colony forming units [CFU]/ml or g) of a specific organism in a given

food. Such applications have been recently proposed for foodborne pathogens such as *Listeria monocytogenes* and *Campylobacter jejuni* (Rantsiou, Alessandria, Urso, Dolci & Cocolin, 2008; Rantsiou, Lamberti & Cocolin, 2010). Most importantly however, qPCR can be used to quantify expression of a target gene. Gene expression can be reported as *comparative* or *absolute*. In the comparative expression case, the expression of a gene is monitored in two different conditions, for example when the microorganism is grown in a synthetic medium and in a real food sample, or is grown at two different temperature conditions. The result is then reported as a ratio and fold change of expression (increase or decrease) between a standard condition (i.e. synthetic medium or high temperature) and an experimental condition (i.e. food matrix or low temperature). To calculate this ratio, it is first necessary to normalize the data, using one or more appropriately chosen, constitutively expressed gene(s) [usually termed housekeeping gene(s)], in order to compensate for potential differences attributed to the preparative steps, mainly the RNA extraction, prior to the qPCR. The selection of genes to be used for normalization is an important aspect of the experimental design and requires validation. This validation entails testing of the expression level, which has to give consistent results, in the different conditions to be analyzed. For absolute quantification, the C_T (threshold cycle) value obtained for the target and normalization genes are transformed into a transcript copy number through calibration curves. In this case, the calibration curves are constructed by plotting C_T values against gene copy numbers, usually obtained by cloning the gene of interest into plasmids

and performing the amplifications with a known quantity (copies) of the plasmid.

Gene expression studies conducted by qPCR so far have proven the versatility of the method (i.e. application in different matrices, use of different primers to study different genes), ease of application (once good quality RNA is extracted from a given matrix, the steps of amplification and data analysis are straightforward) and its capacity for good quality, quantitative data generation. Generally, it is recognized that qPCR is the appropriate method when one needs to study a moderate number of genes in a number of samples that ranges from small to hundreds. On the contrary, *microarrays* offer the possibility for *whole genome* discovery experiments in small number of samples (VanGuilder, Vrana & Freeman, 2008). Furthermore, recently, a trend is being developed towards more function-focused sub-arrays that target specific cellular functions, for example virulence regulons for pathogens or metabolic regulons of interest for technologically important microorganisms. This trend allows application in a larger number of samples and facilitates interpretation of the data obtained. It is important to always keep in mind that data obtained by microarrays, or subarrays (that can be considered of qualitative nature) need to be validated by RT-qPCR.

Statistical treatment of the data and prediction of behavior

Bioinformatics on transcriptomics (and genomics) and chemometrics on metabolomics are applied to take information out of the high-dimensional data produced by the *omics* disciplines. The data may be analyzed by both

unsupervised (use only of *X*-data; see explanation below) such as Hierarchical Clustering (HCA), Principal Component Analysis (PCA), Self-Organizing Maps and Kohonen Neural Networks and supervised methods (use of both *X*- and *Y*-data) such as back-propagation Neural Networks, *k*-nearest neighbors, Discriminant Analysis (DA), Partial Least Square Analysis (PLSA), Partial Least Square Regression (PLSR) and Support Vector Machines. Also, there are special types of supervised methods such as evolutionary-based algorithms, classification and regression trees (CART) and inductive logic programming termed as explanatory or inductive methods (use also of both *X*- and *Y*-data) (Figure 3) (Forina et al., 2004; Tjaden & Cohen, 2006). When learning is unsupervised, the system is shown a set of inputs (*X*-data) and then left to cluster the data into groups. For multivariate analysis this procedure is known as dimensionality reduction. That is, a large amount of data is summarized by fewer parameters with minimal loss of information. When learning is supervised, the desired responses (*Y*-data), associated with each of the inputs (*X*-data), are known. The aim is to find a mathematical transformation (i.e. model) that will correctly associate all or some of the inputs with the target traits. Therefore, the mathematical transformation from input to output data is transparent. Finally, the inductive methods allow the discovery of key inputs for the separation of the traits to be predicted (Forina et al., 2004).

Preliminary statistical analysis of the gene expression matrix

Raw data are usually transformed in expression ratio and fold change. Expression ratio (treated group/control group) constitutes an intuitive

measure of expression changes. Genes without change in their expression have an expression ratio equal to 1. However expression ratio displays some disadvantages relative to up- and down-regulated genes. Expression ratio equal to +2 indicates a change (up-regulation) in the gene expression by a factor of 2 whereas down-regulated genes by the same factor have an expression ratio equal to 0.5. As a consequence, the region of the expression ratios in which the down-regulated genes lie is limited (between 1 and 0) compared to up-regulated genes (between 1 and +infinity). For this reason, the fold change is used instead of expression ratio, which is the reciprocal or the inverse transformation of the expression ratio (when expression ratio ≥1, then fold change = expression ratio; when expression ratio < 1, then fold change = -1 / expression ratio). In this way, a similar representation for the differentially expressed genes is achieved, whether the genes are up- or down- regulated. Therefore, a gene with an increase in its expression by a factor of 2 will have a fold change equal to +2 while a gene with a decrease in its expression by the same factor will have a fold change equal to -2. The use of fold change in the various statistical analyses is problematical only because it is discontinuous between -1 and +1. The best alternative is to apply a logarithmic transformation, generally using the logarithm base 2. The advantage is that it produces a continuous spectrum of values for differentially expressed genes and at the same time treats up- and down-regulated genes equivalently (Causton et al., 2003).

Ratios are good means of comparing levels of gene expression between a treated group and a control group. However, in order to reliably identify

genuine differentially expressed genes an accurate method for comparing the measured expression levels between states should be used. This method is known as normalization, which removes any non-biological variation (artifacts) within the data and allows the data from two samples or states to be appropriately compared. Before proceeding to normalization the data are transformed into $\log_2$ (ratio) values. There are various normalization methods, but the objective of all these methods is all $\log_2$ (ratio) values to be, on average, equal to 0 (Vandesompele et al., 2002; Quackenbush, 2002; Causton et al., 2003; Tjaden & Cohen, 2006).

Advanced statistical analysis of the gene expression matrix

The main goal in such experiments (microarrays and RT-qPCR) is to find genes that are differentially expressed between two states and quantify their expression. However, the appropriate statistical treatment should be applied to the transcriptional data in order to avoid possible errors, which could lead to erroneous conclusions (Yuan, Reed, Chen, & Stewart Jr., 2006; Rebrikov & Trofimov, 2006). After normalizing the data, genes with expression ratios that are significantly different from 1 [or those with values for the $\log_2$ (ratio) that differ from 0] should be identified. The objective of the Analysis of Variance (ANOVA) is to test for significant differences between means by comparing variances. Hence, the variance due to the between-groups (or treatments) variability with that due to the within-group (treatment) variability can be compared (Causton et al., 2003; Tjaden & Cohen, 2006).

The objective of gene clustering is to group together genes that have similar expression profiles. It can be useful for discovering types of behavior or for reducing the dimensionality of the data (e.g. group of genes that behave similarly). Also, clustering helps to identify genes that are co-regulated or that participate in similar biological processes. PCA is one of the most frequently used methods to reduce the dimensionality of data and to find combinations of variables that jointly contribute most to variability in the data. Usually, most of the variability can be accounted for by a small number of principal components (PCs) (two or three). By taking only two or three most important PCs one can visualize the data in two or three dimensions without losing much information (Causton et al., 2003; Tjaden & Cohen, 2006).

The previous two methods (clustering and PCA) try to find structure in the data without using any external information. Classification methods use external information such as annotation and try to find properties in the data that support this information. Therefore, given a gene expression data matrix with samples annotated as 'neutral', 'osmotic' and 'acidic', the classification method will seek for combinations of genes that are expressed in all states. If such genes are found, the knowledge can be used to classify or predict the state of a new unknown sample (Causton et al., 2003). For instance, the objective of the CART analysis shown in Figure 4 is to test if the eighteen initial distinct variables (i.e. genes) of *Listeria monocytogenes* grown under stress (acidic and osmotic) and non stress conditions would allow efficient prediction of the state the stressed cells correspond to and to identify rules that would help classify the stressed

cells on the basis of the measured variables. The classification tree displays the successive steps during which the algorithm identifies the variables that allow the best split of the categories of the dependent variable (Fig. 4). Thus, it can be seen that using the genes *fbpA*, *iap* and *actA*, the algorithm has found rules that allow perfect separation (100% of purity) the various cell states (nodes 4-7). The rules related to 'acidic' and 'neutral' states are verified by all cases (frequency = 9) but for 'osmotic' state there are two different rules to separate all cases indicating potential high variability of gene expression in osmotic condition. The rules that correspond to the leaves of the tree (the terminal nodes) allow for predictions for each observation, with a probability that depends on the distribution of the categories at the leaf level.

The significance of transcriptomics lies in the potential of linking specific changes in gene expression with a phenotype of interest (e.g. stress response). In other words, how expression controls protein production and ultimately the phenotypic characteristics (Yoshida et al. 2001). This will provide insight into the function of various cellular activities of pathogenic bacterial cells at the genetic and transcriptional level. The importance of statistical analysis is to retrieve useful information out of the multivariate data in order to achieve the above objectives.

In situ application of RT-qPCR and microarrays for virulence and stress response determination

The first applications of RT-qPCR and microarrays in the field of food safety and microbiology concerned the study of genes that may play a role in

pathogenicity of microorganisms, *in vitro*. *Listeria monocytogenes* is a foodborne pathogen that currently causes great concern to food industry, authorities and consumers alike, mainly due to its ubiquitous nature, the severe disease it causes and its mechanisms of virulence and adaptation to modern food preservation strategies. As a consequence, *L. monocytogenes* has become one of the first model organisms to study virulence and stress response gene expression (Cossart & Archambaud, 2009), but also others, such as *Escherichia coli* have been investigated (Olesen & Jespersen, 2010). RT-qPCR and to a lesser extent microarrays have been employed in studies *in vitro* (i.e. in laboratory media and under conditions that may influence gene expression) and *in vivo* (i.e. using animals, nematodes or mammalian cell lines to understand the response to eukaryotic intracellular environment). However, only recently the missing link of *in situ* analysis (i.e. analysis of expression in real foods and under common conditions of storage or consumption), is being taken into consideration by food microbiologists. Also in this case, *L. monocytogenes* is among the first foodborne pathogens to be considered. Food is the vehicle through which *L. monocytogenes* enters the human body where, under certain circumstances, it elicits disease. Recent findings have shown that the 'history' of *L. monocytogenes* cells, may influence their virulence potential. It was shown that long-term adaptation to acidic and NaCl stress (such as the ones commonly encountered in foods) increased expression of virulence genes and improved adhesion and invasion to Caco-2 cells (Olesen, Vogensen & Jespersen, 2009). These findings suggest that environmental conditions that *L. monocytogenes* may encounter in foods

could influence its virulence potential. Furthermore, it is expected that modulation of expression of stress response genes, under certain food conditions, may enhance survival and directly or indirectly virulence. *In situ* studies provide valuable data and will complement available information regarding virulence and survival potential leading to an integrated risk assessment analysis.

Liu and Ream (2008) conducted the first *in situ* studies using whole genome microarrays and one strain of *L. monocytogenes* which was implicated in an outbreak of listeriosis and for which the genome sequence was available, thus facilitating analysis of the data. The food taken into consideration was ultrahigh-temperature-processed (UHT) skim milk and the objective was to identify genes whose expression patterns were altered in this matrix. After 24 h incubation at 4 °C, 26 genes were up-regulated in UHT milk compared to Brain Heart Infusion (BHI) broth and 14 were down-regulated. Two genes encoding for proteins involved in oligopeptide uptake by the cell were significantly up-regulated, indicating a possible mechanism for acquisition of essential amino acids. The authors hypothesize that the elevated level of an oligopeptide transport system may result in growth of *L. monocytogenes* in milk. Genes involved in manganese-transport, shown to be related to oxidative stress, were also up-regulated. Also the σ^B encoding gene, playing an important role in both stress response and virulence of *L. monocytogenes* was up-regulated, while expression of other virulence genes was not appreciably altered in UHT milk.

In a more focused study, regarding survival of *L. monocytogenes* on parsley leaves, RT-qPCR was employed to determine expression of 6 genes involved in stress response (*groEL* and *clpC*) and virulence (*bsh*¸ *opuC*, *inlA* and *prfA*) (Rieu, Guzzo & Piveteau, 2009). In this case, *L. monocytogenes* EGDe was inoculated on parsley leaves and incubated 5 hours at 25 °C. Apart for *bsh*, for which no significant difference in expression was observed, all other genes were down regulated between the time of inoculation and the 5 h incubation on parsley leaves. In this study, *in vitro* assessment of the virulence potential, confirmed the down-regulation of *inlA* after the 5 h survival of *L. monocytogenes* on parsley leaves, since a dramatic reduction of adhesion and entry into Caco-2 cells was recorded. However, by *in vivo* experiments using chick embryos, it was shown that the virulence potential of *L. monocytogenes* was recovered.

RT-qPCR was also used to determine the incubation temperature effect on the expression of 4 virulence genes (*hlyA*, *actA*, *inlA* and *prfA*), for two *L. monocyotogenes* strains inoculated in salmon (Duodu, Holst-Jensen, Skjerdal, Cappelier, Pilet & Loncarevic, 2010). Temperatures tested were 4 °C (correct storage temperature) and 20 °C (temperature abuse) while the two strains were chosen based on their virulence (low and high). Significant up-regulation was detected for *hlyA* and *inlA* genes only for the low virulence strain, when incubated at 20 °C. Also in this case, the RT-qPCR data were coupled with experiments of invasion into Caco-2 cells and mouse infection. Overall, authors observed that virulence gene expression, invasion and *in vivo* virulence were not significantly altered under the experimental conditions tested for the highly virulent strain. On the other

hand, up-regulation of 2 out of 4 virulence genes at 20 °C for the low virulence strain was coupled with increased invasion and *in vivo* virulence under the same condition, suggesting an overall increase of its virulence potential at the abuse temperature.

Lastly, the relative transcription of 2 virulence (*prfA* and *inlA*) and 2 stress response (*clpC* and *sigB*) genes was determined in liver pâtés with varying NaCl content (Olesen, Thorsen & Jespersen, 2010). Three different *L. monocytogenes* strains (EGDe, a salt sensitive and a salt resistant strain) were inoculated in liver pâtés with 4 different NaCl concentrations varying from 1.39 to 3.66 % (w/v) NaCl in the water phase (one standard and three reduced) and incubated at 7 °C for 48 h. The liver pâté with the lowest NaCl concentration (i.e. 1.39 % [w/v]) was additionally supplemented with Ca-acetate (0.24 % [w/v]) and Ca-lactate (1.46 % [w/v]) as preservatives. When the standard liver pâté was compared to BHI, no significant differences were observed in the transcription levels for the EGDe strain. On the contrary, *prfA* (for the salt sensitive), *inlA* and *sigB* (for both salt sensitive and resistant) were down-regulated in the liver pâté with respect to the BHI under identical temperature and time conditions. Comparing the standard pâté with the reduced salt pâtés, the differences observed were significant in the case of the *clpC* gene for the EGDe and the salt resistant strain with a significant up-regulation in the pâté with the lowest NaCl concentration, which was also supplemented with Ca-acetate and Ca-lactate. Additionally, significant increase in the transcription was observed for *sigB* when the salt resistant strain was inoculated in pâté with an intermediate NaCl concentration and with low NaCl and Ca-acetate and

Na-acetate. Furthermore, *prfA* showed increased transcription in the pâté with intermediate salt content for the salt sensitive strain. In conclusion, the study showed significant strain variations at the transcriptional level for both stress and virulence genes. Additionally, it was shown that a change in preservation strategy, including lower NaCl content and addition of organic acids as preservatives, can change the transcriptional level of genes related to stress as well as genes controlling the expression of virulence genes.

Experimental challenges

Comparison of the results obtained by the different experimental approaches followed by the authors has revealed some discrepancies. Gene expression studies were coupled to *in vitro* adhesion and invasion studies and *in vivo* virulence determination, using animal models, and the results obtained did not always correlate well. Such differences could partly be expected since the biological basis underlying these experimental approaches is different. An important question though is raised: which approach or combination of approaches is appropriate in order to gather information regarding virulence, potentially to employ in risk assessment? In answering this question, one has to take into consideration also aspects related to the 'applicability' of a certain experimental approach (for example, ethical issues in using animal models, cost of equipment or material).

Challenges still remain in the *in situ* determination of survival and virulence potential. Regarding the methodology, maybe the most critical parameter

still necessitating improvement is the extraction of high quality RNA. This issue is being addressed by different scientists (Monnet, Ulvé, Sarthou & Irlinger, 2008, Ulvé, Monnet, Valence, Fauquant, Falentin & Lortal, 2008, Olesen et al, 2010) and what can be said is that RNA extraction requires optimization, based on the combination of food matrix and microorganism to be studied: no 'universal' protocol exists. Otherwise, technological advancements are constantly improving the quality of the data obtained by both RT-qPCR and microarrays. Concerning the interpretation of the results obtained, as discussed above, there is a need to correlate the gene expression data with valid indicators of survival (i.e. internal pH, membrane integrity) and virulence (i.e. adhesion, invasion) in order to apply more widely such approaches for the determination of these physiological characteristics of foodborne pathogens. Finally, in order to further improve food safety, the perspective use of RT-qPCR and microarrays in conditions that resemble the human gastrointestinal track will fill the scientific information gap regarding the physiology of foodborne pathogens that exists between consumption of contaminated food and disease manifestation. First reports concerning this aspect are available (Jiang, Olesen, Andersen, Weihuan & Jespersen, 2010).

Conclusions and future perspectives

Microbial examination of food products is approaching a new era where additional focus has to be placed on intra-species variations and transcriptional profiling of genes related to the later ability of the pathogens to cause human infections. Based on the knowledge we have

today, plate counting based determination of viable microbial counts is simply not sufficient for proper risk assessment.

The first *in situ* applications of transcriptomics have proven the importance of undertaking approaches that may supply information regarding the potential of foodborne pathogens to survive during food processing and storage, through the passage of the gastrointestinal tract and eventually to cause infection. From results of the studies reported above, two important observations can be made. First, there seems to be significant intra-species heterogeneity concerning stress response and virulence gene expression. It is not yet clear if this heterogeneity is linked to the species of *L. monocytogenes,* but most likely it is valid for other pathogens too. This heterogeneity, furthermore stresses the necessity to move from the determination of numbers of pathogens in a food to the determination of their behavior. Introduction of information regarding the survival and/or virulence potential in the risk assessment evaluation process should be considered. The second observation that can be made relates to the fact that environmental conditions encountered by pathogens during food production, influence gene expression. These two factors (strain and provenience and/or its history during production) need to be taken into consideration in risk assessment, since they influence stress response and virulence potential. Application of methods that allow gene expression determination *in situ* will provide necessary information for risk assessment that considers not only numbers of microorganisms but also their behavior.

Acknowledgements

This review was prepared in the frame of the “Pathogen Combat: control and prevention of emerging and future pathogens at cellular and molecular level throughout the food chain” project, funded by the European Commission within the VI Framework Program, contract n. 007081.

References

Brul, S. (2007). Systems and food microbiology. In S. Brul, S. van Gerwen & M. Zwietering, *Modelling microorganisms in food, 1st edition* (pp. 250-288). Cambridge: Woodhead Publishing Limited.

Causton, H.C., Quackenbush, J. & Brazma, A. (2003). Microarray Gene Expression Data Analysis: A Beginner's Guide. Blackwell Publishing, Oxford, UK.

Cossart, P. & Archambaud, C. (2009). The bacterial pathogen *Listeria monocytogenes*: an emerging model in prokaryotic transcriptomics. *Journal of Biology, 8*(107), 1-4.

Duodu, S., Holst-Jensen, A., Skjerdal, T., Cappelier J.M., Pilet, M.F. & Loncarevic, S. (2010). Influence of storage temperature on gene expression and virulence potential of *Listeria monocytogenes* strains grown in salmon matrix. *Food Microbiology, 27*, 795-801.

Forina, M., Lanteri, S. & Casolino, C. (2004). Chemometrics. In L.M.L. Nollet, *Handbook of Food Analysis: Methods and Instruments in Applied Food Analysis, 2nd edition* (pp. 1757-1804). New York: Marcel Dekker, Inc.

Higuchi, R., Dollinger, G., Walsh, P. S., & Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. *Biotechnology, 10*, 413–417.

Higuchi, R., Fockler, C., Dollinger, G., & Watson, R. (1993). Kinetic PCR: real time monitoring of DNA amplification reactions. *Biotechnology,*

11, 1026–1030.

Jiang, L., Olesen, I., Andersen, T., Weihuan, F. & Jespersen, L. (2010). Survival of *Listeria monocytogenes* in simulated gastrointestinal system and transcriptional profiling of stress- and adhesion-related genes. *Foodborne Pathogens and Disease, 7*, 267-274.

Liu, Y. & Ream, A. (2008). Gene expression profiling of *Listeria monocytogenes* strain F2365 during growth in ultrahigh-temperature-processed skim milk. *Applied and Environmental Microbiology 74*, 6859-6866.

Livak, K.J. & Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_T$ method. *Methods 25*, 402-408.

Monnet, C., Ulvé, V, Sarthou, A.S. & Irlinger, F. (2008). Extraction of RNA from cheese without prior separation of microbial cells. *Applied and Environmental Microbiology, 74*, 5724-5730.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. In *Cold Spring Harbor Symposium on Quantitative Biology. 51,* 263–373.

Olesen, I., Thorsen, L. & Jespersen, L. (2010). Relative transcription of *Listeria monocytogenes* virulence genes in liver pâtés with varying NaCl content. *International Journal of Food Microbiology, 141*, S60-S68.

Olesen, I. & Jespersen, L. (2010). Relative gene transcription and pathogenicity of enterohemorrhagic *Escherichia coli* after long-term

adaptation to acid and salt stress. *International Journal of Food Microbiology 141S*, 248-253.

Olesen, I., Vogensen, F.K. & Jespersen, L. (2009). Gene transcription and virulence potential of *Listeria monocytogenes* strains after exposure to acidic and NaCl stress. *Foodborne Pathogens and Disease 6*, 669-680.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research 29*, 2002-2007.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics, 32, Supplement*, 496-501.

Rantsiou, K., Alessandria, V., Urso, R., Dolci, P., & Cocolin, L. (2008). Detection, quantification and vitality of *Listeria monocytogenes* in food as detennined by quantitative PCR. *International Journal of Food Microbiology, 121*, 99-105.

Rantsiou, K., Lamberti, C., & Cocolin, L. (2010). Survey of *Campylobacter jejuni* in retail chicken meat products by application of a quantitative PCR protocol. *International Journal of Food Microbiology, 141S,* S75-S79.

Rebrikov, D.V. & Trofimov, D.Y. (2006). Real-time PCR: a review of approaches to data analysis. *Applied Biochemistry and Microbiology, 42*, 455-463.

Rieu, A., Guzzo, J. & Piveteau, P. (2009). Sensitivity to acetic acid, ability to colonize abiotic surfaces and virulence potential of *Listeria monocytogenes* EGD-e after incubation on parsley leaves. *Journal of Applied Microbiology 108*, 560-570.

Skovgaard, N. (2007). New trends in emerging pathogens. *International Journal of Food Microbiology, 120*, 217-224.

Tjaden, B. & Cohen, J. (2006). A survey of computational methods used in microarray data interpretation. *Applied Mycology and Biotechnology 6,* 161-178.

Ulvé, V.M., Monnet, C., Valence, F., Fauquant, J., Falentin, H. & Lortal, S. (2008). RNA extraction from cheese for analysis of *in situ* gene expression of *Lactococcus lactis*. *Journal of Applied Microbiology, 105*, 1327-1333.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology 3*, 1-11.

VanGuilder, H.J., Vrana, K.E. & Freeman, W.M. (2008). Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques, 44*, 619-626.

Wilson, C.H., Tsykin, A., Wilkinson, C.R. & Abbott, C.A. (2006). Experimental design and analysis of microarray data. *Applied Mycology and Biotechnology 6*, 1-36.

Yoshida, K., Kobayashi, K., Miwa, Y., Kang, C.M., Matsunaga, M., Yamaguchi, H., Tojo, S., Yamamoto, M., Nishi, R., Ogasawara, N., Nakayama, T. & Fujita, Y. (2001). Combined transcriptome and proteome analysis as a powerful approach to study genes under glucose repression in *Bacillus subtilis*. *Nucleic Acids Research 29*, 683-692.

Yuan, J.S., Reed, A., Chen, F. & Stewart Jr., C.N. (2006). Statistical analysis of real-time PCR data. *BMC Bioinformatics 7*, 1-12.

Legends to Figures

Figure 1. Work-flow of the application of *omics* disciplines. *Omics,* for example, may provide information relative to genes being regulated during microorganisms resistance against the antimicrobial agents that are used in the food industry. This could lead to new hypothesis regarding the microorganisms behavior in a food product; however validation experiments to investigate the interactions between microorganisms and food matrix should be performed to validate the hypothesis made.

Figure 2. Graphical representation of the molecular approach to study the expression profile of genes of pathogenic bacteria isolated from foods.

Figure 3. Analysis of high-dimensional data produced by *omics* disciplines.

Figure 4. Classification tree diagram. Rules found by the algorithm to differentiate the various states of the cell. For instance, if *iap* in [-0.944, -0.228] and *fbpA* in [-1.06, -0.22] then condition = acidic in 100% of cases or if *actA* in [-1.244, 1.947] and *fbpA* in [-0.22, 1.282] then condition = neutral in 100% of cases.

Figure 1

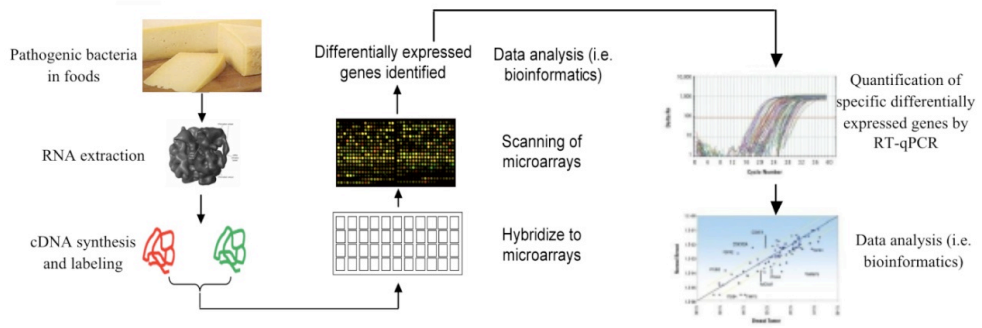


Figure 2

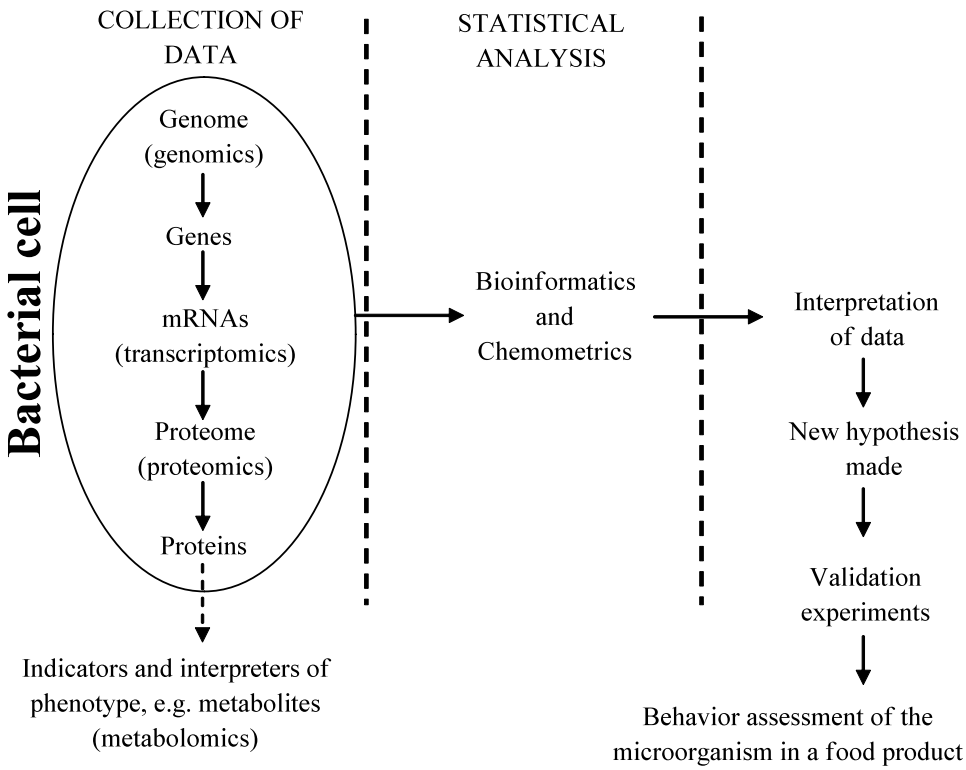


Figure 3

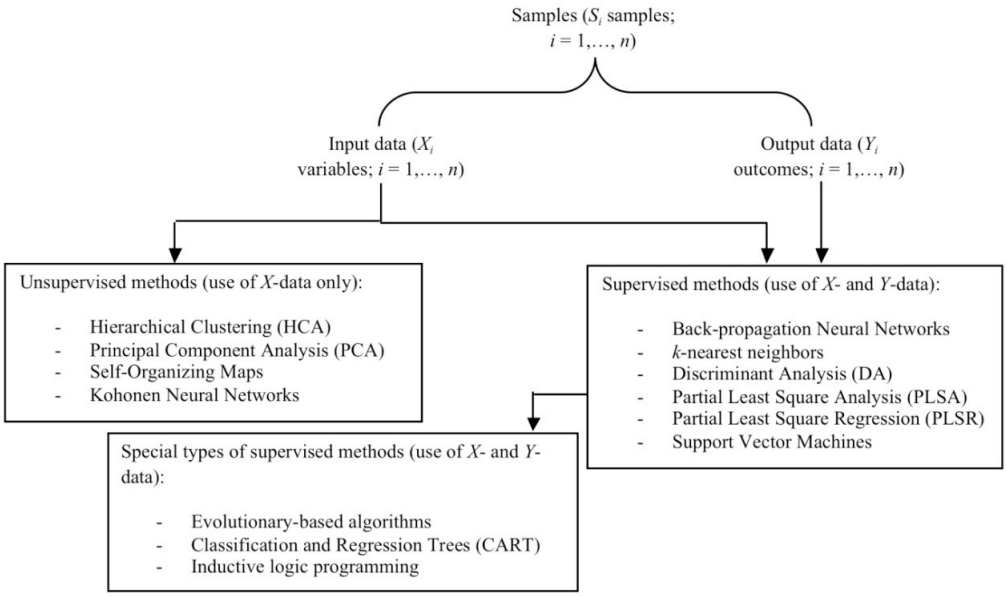
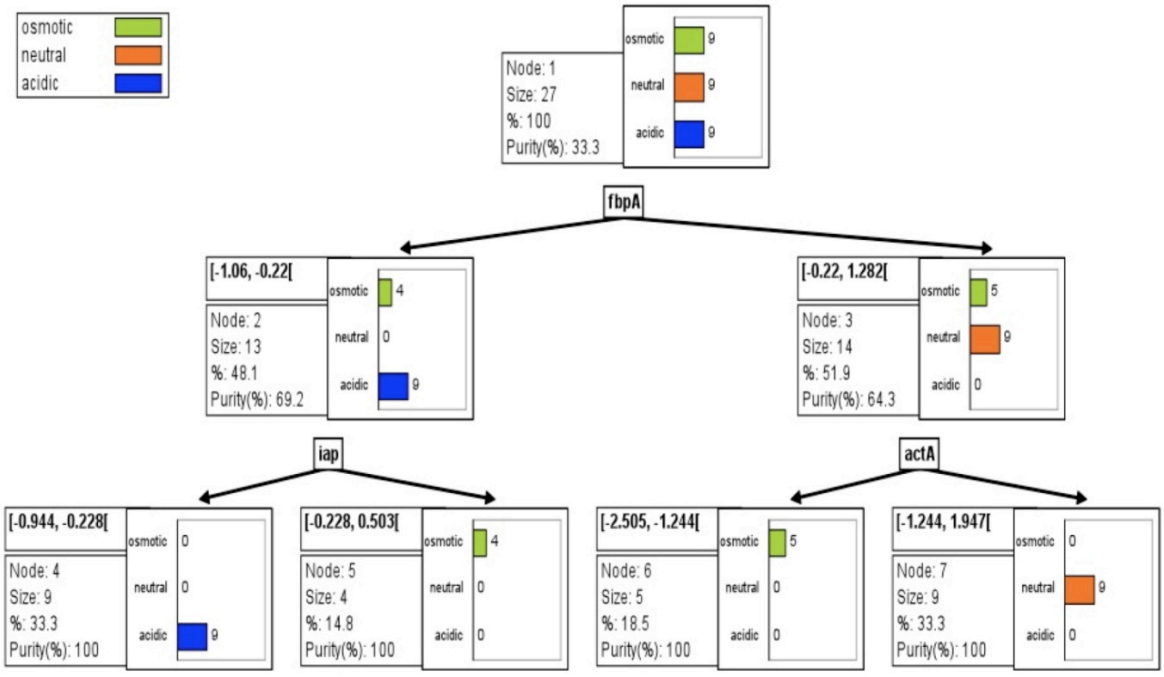

Samples (S_i samples; i = 1,…, n)
Input data (X_i variables; i = 1,…, n)
Output data (Y_i outcomes; i = 1,…, n)
Unsupervised methods (use of X-data only):
- Hierarchical Clustering (HCA)
- Principal Component Analysis (PCA)
- Self-Organizing Maps
- Kohonen Neural Networks
Supervised methods (use of X- and Y-data):
- Back-propagation Neural Networks
- k-nearest neighbors
- Discriminant Analysis (DA)
- Partial Least Square Analysis (PLSA)
- Partial Least Square Regression (PLSR)
- Support Vector Machines
Special types of supervised methods (use of X- and Y-data):
- Evolutionary-based algorithms
- Classification and Regression Trees (CART)
- Inductive logic programming

Figure 4


osmotic
neutral
acidic
Node: 1
Size: 27
%: 100
Purity(%): 33.3
9
9
9
fbpA
[-1.06, -0.22[
Node: 2
Size: 13
%: 48.1
Purity(%): 69.2
4
0
9
[-0.22, 1.282[
Node: 3
Size: 14
%: 51.9
Purity(%): 64.3
5
9
0
iap
actA
[-0.944, -0.228[
Node: 4
Size: 9
%: 33.3
Purity(%): 100
0
0
9
[-0.228, 0.503[
Node: 5
Size: 4
%: 14.8
Purity(%): 100
4
0
0
[-2.505, -1.244[
Node: 6
Size: 5
%: 18.5
Purity(%): 100
5
0
0
[-1.244, 1.947[
Node: 7
Size: 9
%: 33.3
Purity(%): 100
0
9
0